Chapitre I - Statistiques descriptives

I. Rappel de cours

1. Moyenne, écart-type, variance, écart-type estimé, mode, médiane

a. Moyenne, écart-type et variance

Soit un échantillon de n valeurs x. x_i correspond à la $i^{ème}$ valeur de l'échantillon. Soit m la moyenne de l'échantillon, σ_0 son écart-type et σ_0^2 sa variance.

$$m = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \sigma_0 = \sqrt{\frac{\sum_{i=1}^{n} (x_i - m)^2}{n}} \quad \text{ou} \quad \sigma_0 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - m^2}$$

$$\sigma_0^2 = \frac{\sum_{i=1}^{n} (x_i - m)^2}{n} \quad \text{ou} \quad \sigma_0^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - m^2$$

Soit un échantillon de n valeurs x. Il y a k valeurs x_i, correspondant à la $i^{ème}$ valeur de l'échantillon avec un effectif n_i. Soit m la moyenne de l'échantillon, σ_0 son écart-type et σ_0^2 sa variance.

$$m = \frac{\sum_{i=1}^{k} n_i\, x_i}{n} \qquad \sigma_0 = \sqrt{\frac{\sum_{i=1}^{k} n_i\, (x_i - m)^2}{n}} \quad \text{ou} \quad \sigma_0 = \sqrt{\frac{\sum_{i=1}^{k} n_i\, x_i^2}{n} - m^2}$$

$$\sigma_0^2 = \frac{\sum_{i=1}^{k} n_i\, (x_i - m)^2}{n} \quad \text{ou} \quad \sigma_0^2 = \frac{\sum_{i=1}^{k} n_i\, x_i^2}{n} - m^2$$

On parlera également de la fréquence relative f_i avec $f_i = \frac{n_i}{n}$

Dans ce contexte,

$$m = \sum_{i=1}^{k} f_i x_i \qquad \sigma_0 = \sqrt{\sum_{i=1}^{k} f_i (x_i - m)^2} \quad \text{ou} \quad \sigma_0 = \sqrt{\left(\sum_{i=1}^{k} f_i x_i^2\right) - m^2}$$

$$\sigma_0^2 = \sum_{i=1}^{k} f_i (x_i - m)^2 \quad \text{ou} \quad \sigma_0^2 = \left(\sum_{i=1}^{k} f_i\, x_i^2\right) - m^2$$

Remarques

Soit une série de k fréquences relatives f_i pour un échantillon. f_i est la $i^{ème}$ valeur de la fréquence relative de l'échantillon.

$$\sum_{i=1}^{k} f_i = 1$$

b. Mode et médiane

Soit un échantillon de n valeurs x. Il y a k valeurs x_i, correspondant à la $i^{ème}$ valeur de l'échantillon avec une fréquence relative f_i. Le mode est la valeur de x_i ayant la plus grande fréquence relative f_i.

Soit un échantillon de n valeurs x. x_i correspond à la $i^{ème}$ valeur de l'échantillon.
Si n est impair, la médiane est la valeur de x_i telle que $n_1 = n_2$, n_1 étant le nombre de valeur de x_1 à x_{i-1} et n_2 le nombre de valeur de x_{i+1} à x_n.

Si n est pair, la médiane est la valeur $x = \dfrac{x_i + x_{i+1}}{2}$ telle que $n_1 = n_2$, n_1 étant le nombre de valeur de x_1 à x_i et n_2 le nombre de valeur de x_{i+1} à x_n.

c. Ecart-type estimé

Soit un échantillon de n valeurs x. x_i correspond à la $i^{ème}$ valeur de l'échantillon. Soit σ_0 son écart-type. Si cet échantillon provient d'une population inconnue, alors on pourra estimer l'écart-type de cette population, que l'on appellera s.

$$\frac{s^2}{n} = \frac{\sigma_0^2}{n - 1}$$

2. Propriétés de la moyenne et de la variance

Soit un échantillon X. La moyenne peut s'écrire $\overline{X}$ et la variance V(X).

$\overline{aX} = a\,\overline{X}$ $\qquad \overline{X + Y} = \overline{X} + \overline{Y}$

$V(X) = \overline{X^2} - \left(\overline{X}\right)^2 \qquad V(aX) = a^2\, V(X) \qquad V(X + b) = V(X)$

3. Changement de variable

Soit un échantillon X et on considère un échantillon X' dont on connait la moyenne $m_{X'}$ et l'écart-type σ_0', lié à X par le changement de variable $X' = \dfrac{X - X_0}{h}$ (X_0 et h sont connus et $h \neq 0$). Soit m_X la moyenne de l'échantillon X et son écart-type σ_0.

$$m_X = X_0 + h\, m_{X'} \qquad\qquad \sigma_0 = h\,\sigma_0'$$

4. Répartition par classe ou en cumulation

Lorsqu'un nombre de données est important ou lorsque l'on travaille avec des variables continues, on utilisera la notion de classe pour ranger ces données. La classe commencera avec une borne minimale x_{1min} et s'arrêtera à une borne maximale x_{2min} non incluse, qui est la borne minimale de la classe supérieure. Ainsi on aura toute une série de classes et elles devront avoir le même intervalle :

$$x_{2min} - x_{1min} = x_{3min} - x_{2min} = x_{4min} - x_{3min} = x_{5min} - x_{4min} = \text{etc.}$$

Il sera affecté à chacune de ces classes un effectif relatif n_i ou alors une fréquence relative f_i.

On pourra également retrouver des répartitions en cumulation, avec des effectifs cumulés et des fréquences cumulées. On ne parlera plus de répartition par classe. Il sera considéré pour la première valeur de fréquence cumulée toutes les valeurs de x telles que $x < x_{2min}$. Pour la deuxième valeur de fréquence cumulée, on considèrera toutes les valeurs de x telles que $x < x_{3min}$ (incluant les valeurs $x < x_{2min}$).

Pour le calcul des paramètres statistiques, une classe sera définie par une seule valeur x_i telle que $x_i = \frac{x_{(i+1)\min} - x_{(i)\min}}{2}$

4. Utilisation de la calculatrice

Lors du concours, la calculatrice officielle du concours pourra être utilisée et elle possède une fonctionnalité « statistique ». Ainsi, il sera possible de déterminer directement la moyenne, l'écart-type et l'écart-type estimé pour un échantillon.

II. QCM série I

Série de QCM illustrant le chapitre I sur les statistiques descriptives.

QCM SI-1. **Soit un échantillon de n valeurs x. x_i correspond à la $i^{ème}$ valeur de l'échantillon. Soit m la moyenne de l'échantillon et σ_0 son écart-type. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A $m = \sqrt{\dfrac{\sum_{i=1}^{n} x_i}{n}}$

B $m = \dfrac{\sum_{i=1}^{n} x_i}{n}$

C $\sigma_0 = \sqrt{\sum_{i=1}^{n} n\,(x_i - m)^2}$

D $\sigma_0 = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - m)^2}{n}}$

E $\sigma_0^{\,2} = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - m)^2}{n}}$

QCM SI-2. **Soit un échantillon de n valeurs x. Il y a k valeurs x_i, correspondant à la $i^{ème}$ valeur de l'échantillon avec un effectif n_i. Soit m la moyenne de l'échantillon et σ_0 son écart-type. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A $m = \dfrac{\sum_{i=1}^{k} n_i \; x_i}{n}$

B $m = \dfrac{\sum_{i=1}^{k} x_i}{f_i}$ avec $f_i = \dfrac{n_i}{n}$

C $\sigma_0 = \sqrt{\dfrac{\sum_{i=1}^{k} (n_i \; x_i - m)^2}{n}}$

D $\sigma_0 = \sqrt{\sum_{i=1}^{k} n_i \; (x_i - m)^2}$

E $\sigma_0{}^2 = \dfrac{\sum_{i=1}^{k} f_i \; x_i^2}{n} - m^2$ avec $f_i = \dfrac{n_i}{n}$

QCM SI-3. **Soit un échantillon de n valeurs x. Son écart-type est σ_0 et l'écart-type estimé de la population inconnue d'où provient l'échantillon est s. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A $s = \sigma_0 \sqrt{\dfrac{n}{n-1}}$

B $s = \sigma_0 \sqrt{\dfrac{n-1}{n}}$

C $\sigma_0{}^2 = s^2 \dfrac{n}{n-1}$

D $\sigma_0{}^2 = s^2 \dfrac{n-1}{n}$

E Aucune des propositions A, B, C et D précédentes n'est correcte.

QCM SI-4. **La concentration d'une molécule dans le sang est mesurée sur un échantillon de personnes. On obtient les résultats suivants (en ng/L): 17,5 ; 12,5 ; 13 ; 14,5 et 11,5. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A L'écart-type de l'échantillon est $\sigma_0 = 2,33$ ng/L
B L'écart-type de l'échantillon est s = 2,33 ng/L ;
C Sachant que la population n'est pas connue, l'écart-type de la population représentative de l'échantillon est $\sigma_0 = 2,33$ ng/L
D Sachant que la population n'est pas connue, l'écart-type de la population représentative de l'échantillon est s = 2,33 ng/L
E La variance de l'échantillon est $\sigma_0^2 = 4,36$ ng/L

QCM SI-5. **Lors d'une session d'examen, 16% des étudiants ont eu la note de 25, 54 % la note de 50, et 30% la note de 75. On considère m la moyenne de l'échantillon, σ_0 son écart-type et s l'écart-type de la population inconnue d'où provient l'échantillon. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A Il est possible de calculer m
B Il est possible de calculer m et σ_0
C Il est possible de calculer m, σ_0 et s
D Il est impossible de calculer σ_0 car la taille n de l'échantillon n'est pas connue
E Il est impossible de calculer s car la taille n de l'échantillon n'est pas connue

QCM SI-6. **On se propose de mesurer la concentration d'une molécule active sur un échantillon de X flacons. x_i correspond à la concentration du i^ème^ flacon de l'échantillon X. On trouve $\sum_{i=1}^{10} x_i = 300$ mg/L et $\sum_{i=1}^{10} (x_i)^2 = 20000$ mg²/L². On considère m la moyenne de cet échantillon et σ_0 son écart-type. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A m = 3000 mg/L et $\sigma_0 = 1100$ mg/L
B m = 3000 mg/L et $\sigma_0^2 = 1100$ mg²/L²
C m = 30 mg/L et $\sigma_0 = 33,17$ mg/L
D m = 30 mg/L et $\sigma_0 = 1100$ mg/L
E m = 30 mg/L et $\sigma_0^2 = 1100$ mg²/L²

QCM SI-7. **Après avoir effectué un changement de variable $X' = \dfrac{X - X_0}{h}$ avec X_0 = 200000 et h = 500, on trouve que la moyenne des valeurs X' est m_{X}' = 10 et son écart-type σ_0' = 0,1. Soit m_X la moyenne des valeurs X et σ_0 son écart-type. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).**

A $m_X = 195000$ et $\sigma_0 = 50$
B $m_X = 195000$ et $\sigma_0 = 20000$
C $m_X = 205000$ et $\sigma_0 = 50$
D $m_X = 205000$ et $\sigma_0 = 20000$
E Aucune des réponses A, B, C et D n'est correcte

QCM SI-8. **On a relevé l'âge des étudiants d'une promotion PACES. Les résultats ont été répartis par classe avec les fréquences correspondantes dans le tableau suivant.**

Classe des âges (an)	Fréquences
16 – 17	0,002
17 – 18	0,007
18 – 19	0,283
19 – 20	0,455
20 – 21	0,198
21 – 22	0,054
22 – 23	0,001
Nombre total d'étudiants de la promotion	1100

Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).

A La moyenne de l'échantillon est 20,51 ans
B L'écart-type de l'échantillon est 0,97 ans
C La variance de l'échantillon est 0,75 ans^2
D L'écart-type estimé de l'âge de tous les étudiants de promotion PACES en France (population inconnue) est 0,87 ans
E La médiane de l'échantillon est 0,455

QCM SI-9. On a relevé la masse des étudiants d'une promotion PACES. Les résultats ont été répartis par classe avec les effectifs cumulés correspondant dans le tableau suivant.

Classe des masses (kg)	Effectifs cumulés
Moins de 66	**20**
Moins de 69	**190**
Moins de 72	**480**
Moins de 75	**890**
Moins de 78	**1050**
Nombre total d'étudiants de la promotion	**1100**

On considère m la moyenne de cet échantillon et σ_0 son écart-type. Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s).

A L'échantillon est composé de 5 classes
B Le mode de l'échantillon est 75 kg
C m = 70,33 kg et σ_0 = 2,31 kg
D m = 72,33 kg et σ_0 = 3,31 kg
E m = 74,33 kg et σ_0 = 4,31 kg